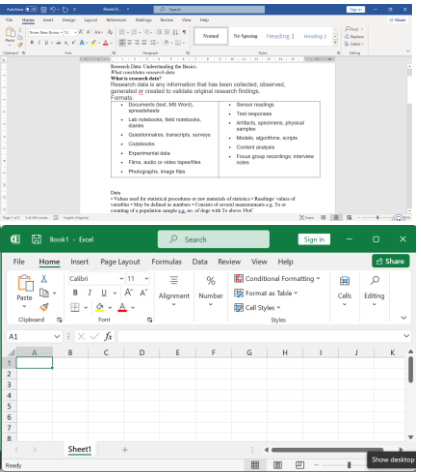
Research Data: Understanding the Basics

Ismail A. Odetokun
DVM, MVPH, Ph.D., FCVSN, MNYA
Associate Professor
Department of Veterinary Public Health & Preventive Medicine
University of Ilorin, Nigeria
odetokun.ia@unilorin.edu.ng; +234 (0)803 088 6018

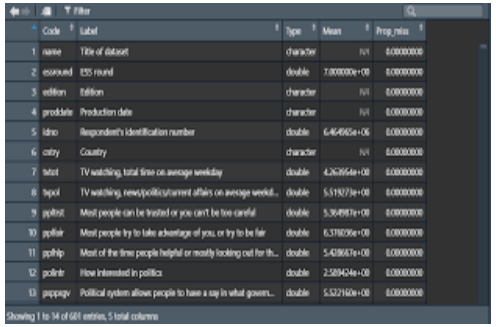
What constitutes research data?

Lab notebooks, field notebooks, diaries

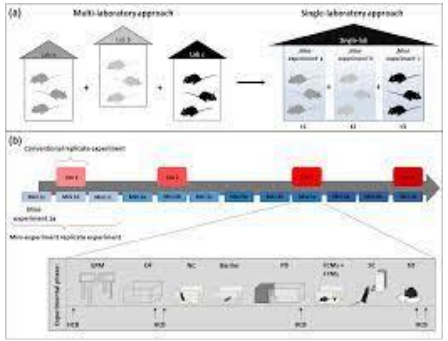
Questionnaires, transcripts, surveys



Documents (text, MS Word), spreadsheets



Codebooks

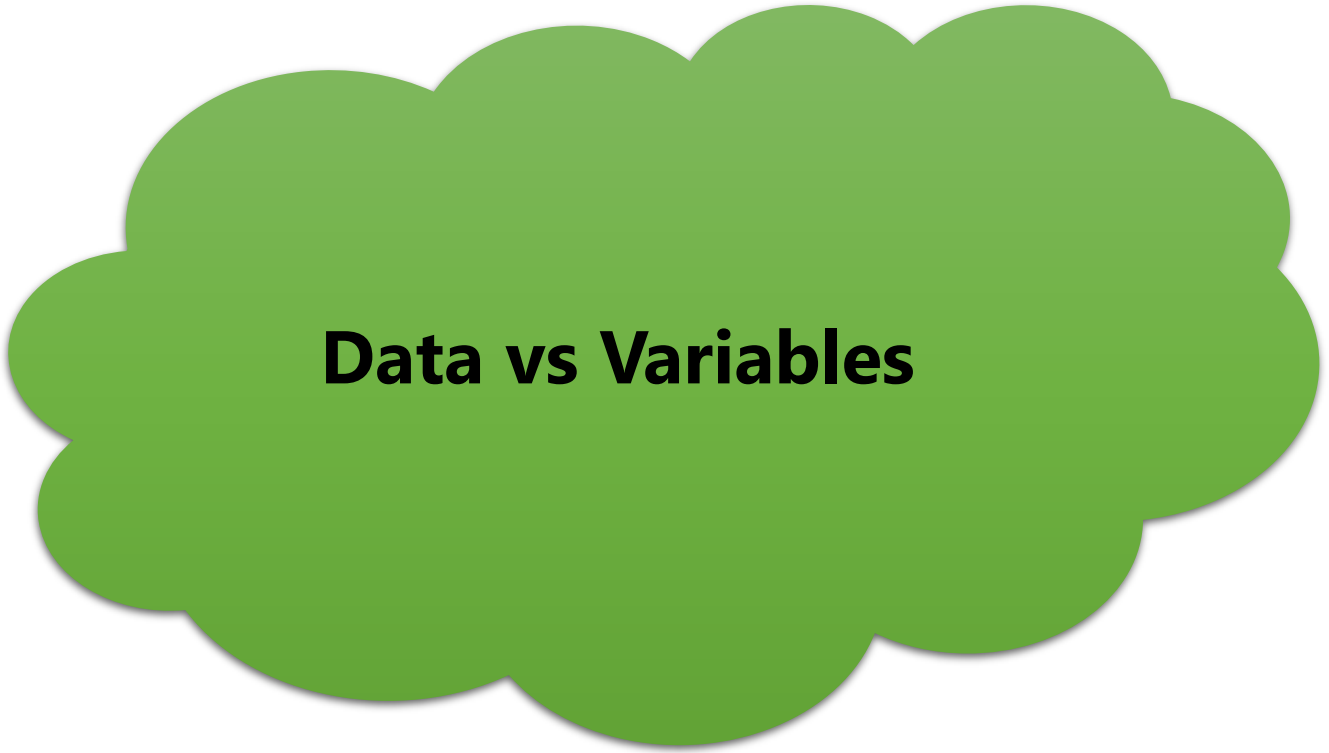


Experimental data

Research data is any information that has been collected, observed, generated, or created to validate original research findings

- **Sensor readings**
- **Test responses**
- **Artifacts, specimens, physical samples**
- **Models, algorithms, scripts**
- **Content analysis**
- **Focus group recordings; interview notes**

Films, audio or video tapes/files


Data vs Variables

Data

Values used for statistical procedures or raw materials of statistics

Readings/values of variables

May be defined as numbers

Consists of several measurements

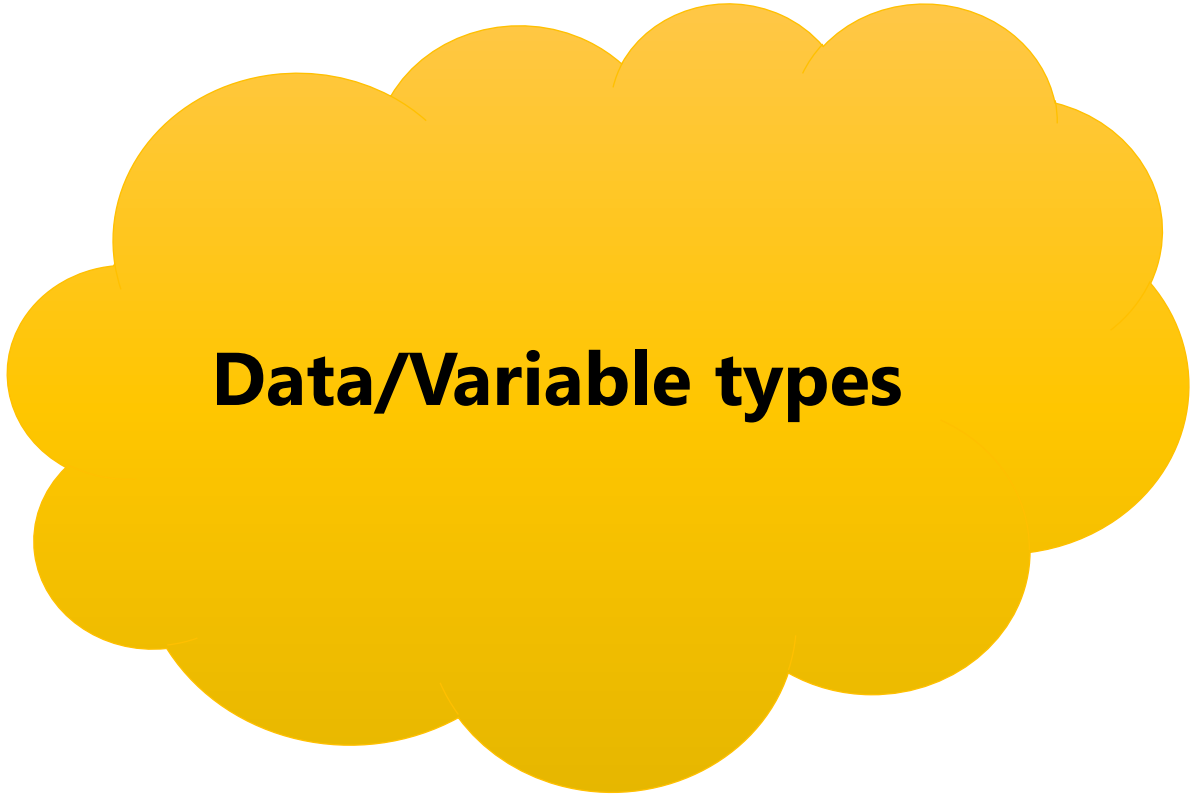
Variables

characteristics that can be measured and that can assume different values

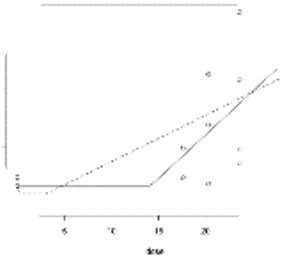
can attain different values from person-to-person, place-to-place or time-to-time

examples

- **height**
- **weight**
- **age**
- **heart rates**
- **litter size**
- **blood count,**
- **enzyme activity**
- **coat colour**
- **percentage of a flock that is pregnant**
- **temperature, etc**

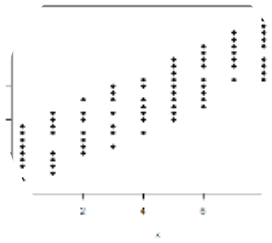

Data/Variable types

Data

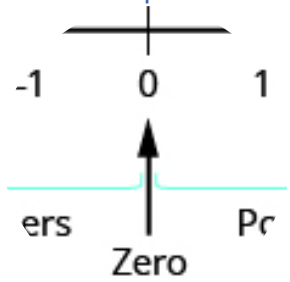


Quantitative

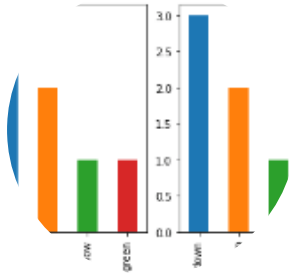
Qualitative



Discrete



Continuous



Categorical

Qualitative/Categorical

readings of characteristics that cannot be measured but labelled

non-numerically valued [no. and percentage (proportion)]

breed

Sex

Colour

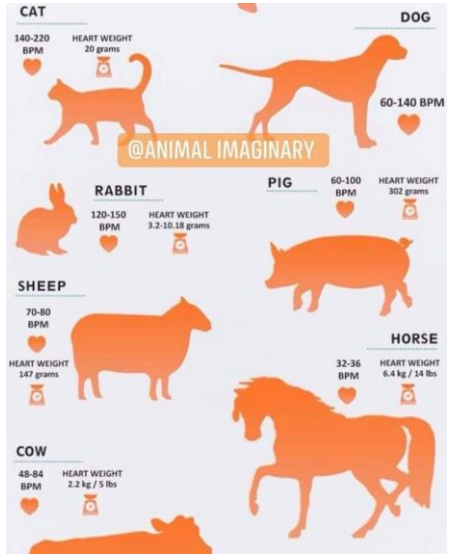
- Name
- Nationality
- States
- Occupation
- Place of birth
- Disease prognosis
- Stages of cancer
- Degree of pain, etc

Quantitative

numerically valued

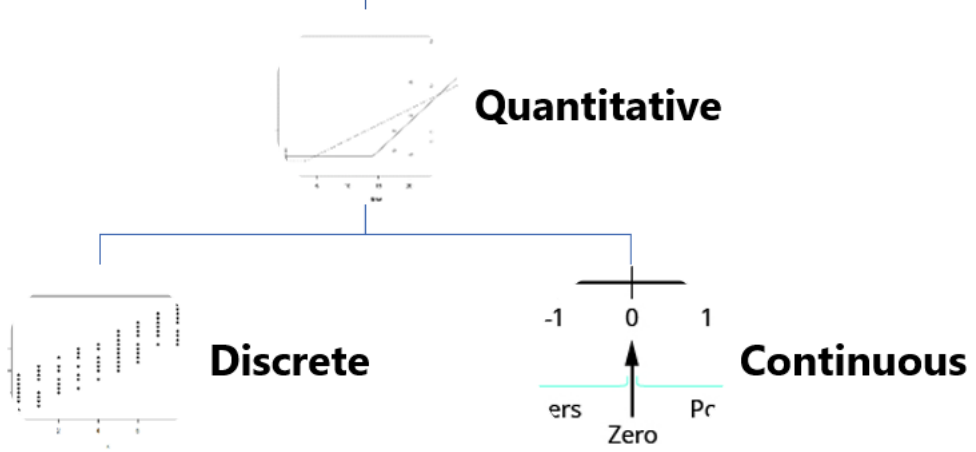
one that can be measured

Height



Weight

- Age
- Heart rate
- Pulse rate
- Temperature
- Water volume
- Density
- Speed
- Force, etc



Discrete

Whole

- **Number of piglets or puppies**
- **Parity (no. of pregnancies)**
- **Episodes of diarrhoea**

Continuous

- **height,**
- **weight,**
- **skull circumference**
- **1m,**
- **1.2m,**
- **50kg,**
- **50.5kg,**
- **51kg**
- **etc**

Data Types

Measurement Scales

There are four types of measurement scales used in statistics: nominal, ordinal, interval and ratio. Each scale has different properties and uses.

QUALITYGURUS

1

Nominal

The category labels are not ordered, so it doesn't matter which number comes first.

No more or less

2

Ordinal

In the ordinal scale of data, there is an order. However, the difference between them can not be quantified.

More or less (not how much)

3

Interval

In the interval scale, we do have an order (just like ordinal data), and we can find the exact difference between the two values.

How much more or less

4

Ratio

The ratio scale has all the features of the Interval scale, and in addition, there is an absolute or true zero as well.

1,2 = Categorical while 3 & 4 = Quantitative


Sources of data

Primary sources (prospective): collected directly from the data source without going through any existing source

- Surveys
- Experiments
- Questionnaires
- Observing interventions
- Focus groups
- Interviews etc

Secondary sources (retrospective): collected in the past by someone else but made available for others to use.

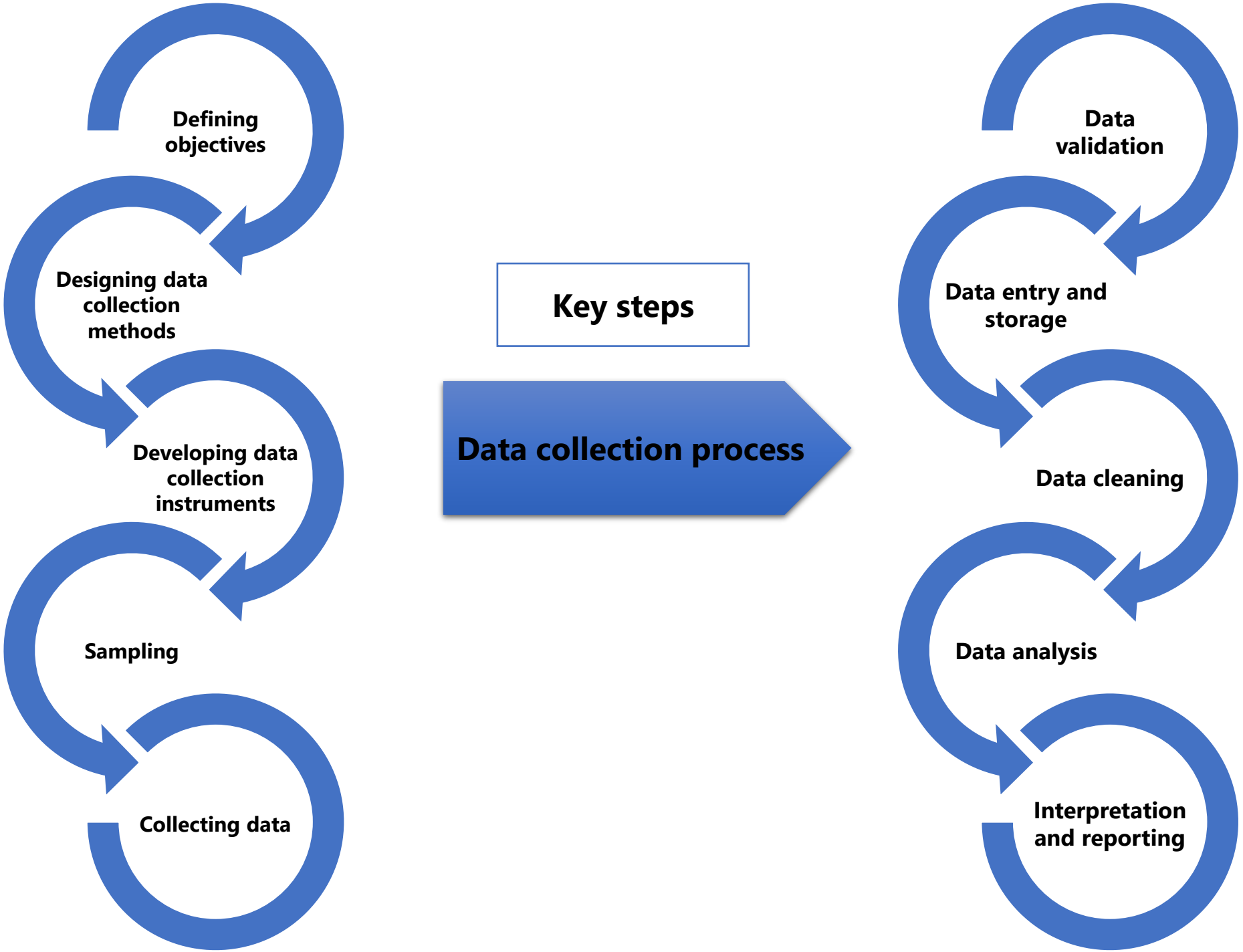
- Routinely kept records
- External sources
- Such as from the library, books, journals, articles, web pages bots, abattoir records, clinic records, hospital records, census data, NBS, etc.

Data collection

the process of gathering and measuring information on targeted variables in a systematic and organized manner.

it involves the acquisition of data from various sources

Purpose: to obtain accurate, reliable, and relevant data for analysis, decision-making, research, etc.


Defining objectives
Designing data collection methods
Developing data collection instruments
Sampling
Collecting data
Key steps
Data collection process
Data validation
Data entry and storage
Data cleaning
Data analysis
Interpretation and reporting

Data collection tools

Surveys and Questionnaires

administered in person, via mail, email, online forms, or through mobile apps.

SurveyMonkey

qualtrics XM

Google Forms
Software

Typeform

Data collection tools

Interviews

direct interaction between the interviewer and the interviewee to gather qualitative data (structured, semi-structured, or unstructured)

Specialized interview management platforms

Data collection tools

Observations

systematically watching and recording behaviors, events, or phenomena

Data collection tools

Experiments

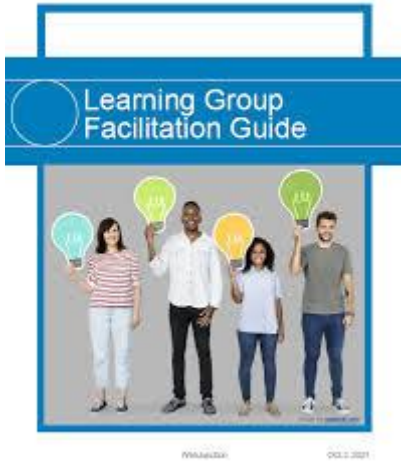
manipulating variables under controlled conditions to observe their effects on outcomes of interest

Data collection tools

Focus Groups

involve facilitated group discussions with a small number of participants to gather insights and opinions on a specific topic

Transcription software

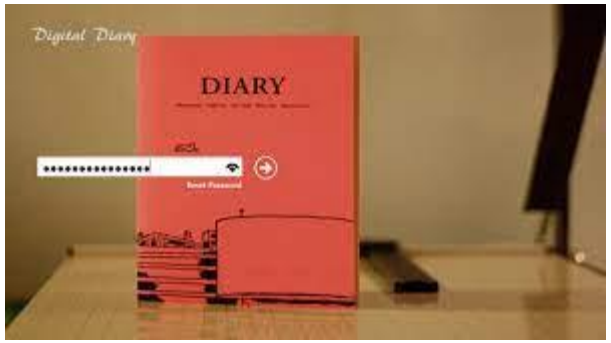


Audio and video recording equipment

Data collection tools

Diaries and Logs

involve facilitated group discussions with a small number of participants to gather insights and opinions on a specific topic



- **paper diaries**
- **digital diary apps**
- **online platforms for data entry**

Data collection tools

Social Media Monitoring

Social media platforms provide a wealth of data that can be collected and analyzed for research purposes

Hootsuite

sproutsocial

social listening platforms

Brandwatch

Data collection tools

Sensor Data Collection

Sensors can collect various types of data, such as temperature, humidity, motion, or physiological signals

- Level sensors
- Chemical sensors
- Proximity sensors
- Humidity sensors
- IoT sensor connectivity

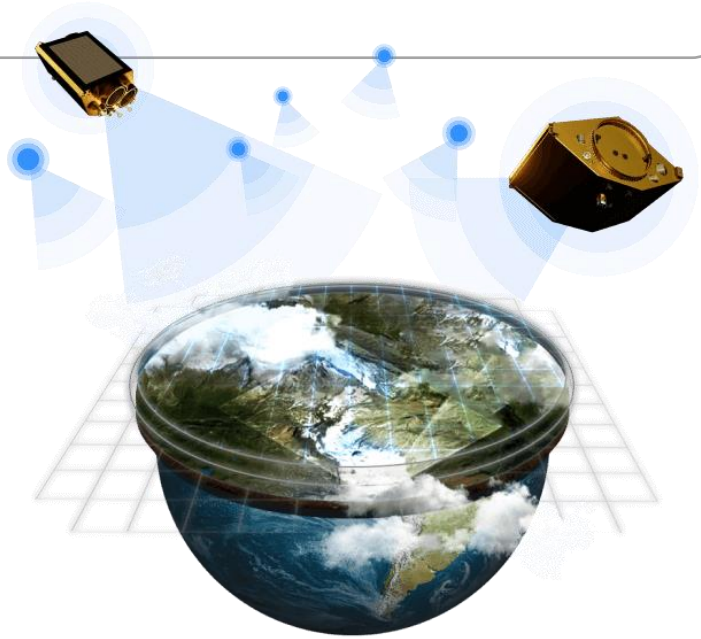
- Temperature sensors
- Infrared sensor
- Biomedical sensors
- Cameras
- Leak sensors
- Accelerometer
- Pressure
- Gyroscope
- Heart rate

- **sensor devices**
- **data loggers**
- **software for data visualization and analysis**

Data collection tools

Geospatial Data Collection

involves capturing and analyzing spatially referenced data, such as GPS coordinates or GIS data



- **GPS devices**
- **GIS software**
- **remote sensing technologies**

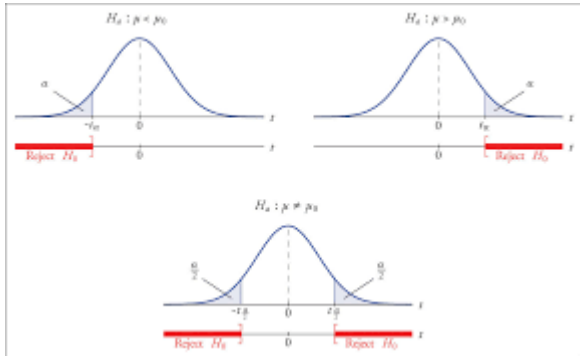
QGIS

ArcGIS

MapInfo

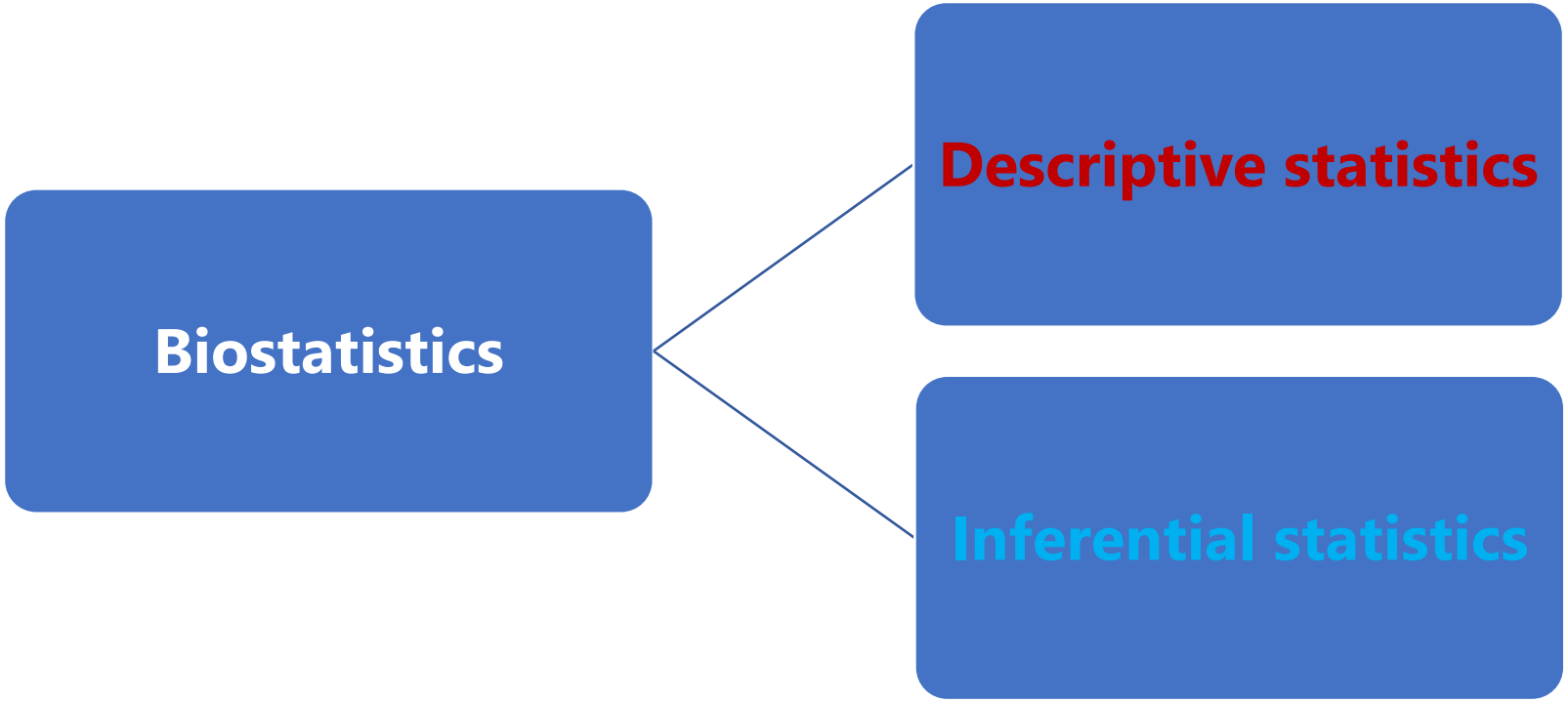

What are the usefulness of good knowledge of data in research?

- **Effective Analysis**
- **Data Interpretation**
- **Hypothesis Testing**
- **Reproducibility and Transparency**
- **Data-driven Decision Making**
- **Quality Assurance**



Data in biostatistics

Is the scientific method of collecting, organizing, presenting, summarizing, analyzing, and interpreting medical and research data



Tools in Descriptive Statistics

Frequency tables

Percentage, frequency distribution, relative frequency, & cumulative frequency

Diagrams, Charts or Graphs

Qualitative:
Pie chart and bar chart
Quantitative:
Histogram, frequency polygon, stem and leaf plot, box plot, scatter diagram

Summary statistics or indices

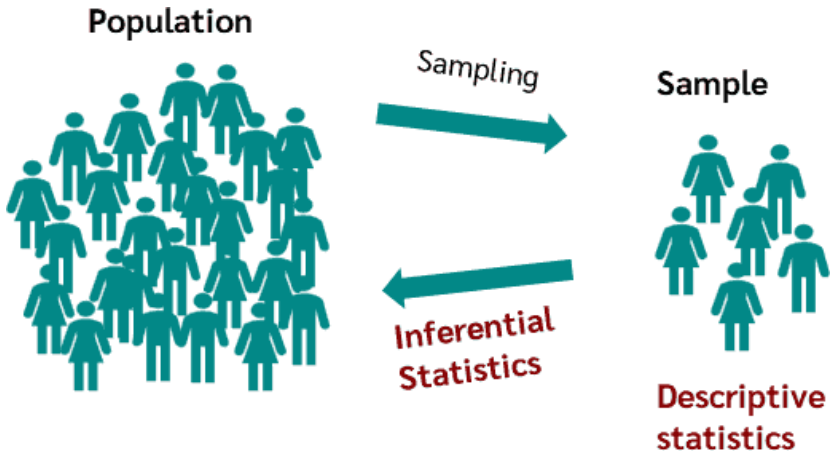
Location/Central tendency:
Arithmetic mean, median, mode, geometric mean, etc
Dispersion/spread:
Range, variance, standard deviation, etc

Inferential statistics

Probability distributions

Hypothesis testing

Confidence intervals



Parametric tests

- One-sample t-test
- Two-sample t-test (Independent)
- Two sample t-test (dependent)
- Correlation/Regression
- One-way ANOVA
- One-way blocked ANOVA or two-way ANOVA

Non parametric tests

- Sign Test
- Mann-Whitney Test
- Wilcoxinsign ranked Test
- Spearman Rank Test
- Kruskal-Wallis Test
- Friedman Test

Choice of test statistics

Depends on

- **Study objective** (estimation, comparison)
- **Study design** (independent or dependent samples)
- **Variable (data) type** – quantitative or qualitative
- **Sample size and sampling method** (random or not)
- **Sampling distribution**

Selecting the right test (>1 samples)

Relationship between 2 quantitative variables

- Correlation analysis(strength of relationship)
- Linear regression analysis

Relationship between 2 qualitative variables

- Chi square test

Relationship between one qualitative and one quantitative variable

- Can be reduced to mean difference between two or more groups
- if two groups: t-test
- If >2 groups: Analysis of variance (ANOVA)

Statistical software tools for researchers


THANK YOU